Error Propagation and Statistical Validation of Computer Vision Software

Xufei Liu,[1] Tapas Kanungo,[2] and Robert M. Haralick[3]

[1]Ranch Networks
Moganville, NJ 07751

[2]Center for Automation Research
University of Maryland, College Park, MD 20742-3275

[3]Dept. of Electrical Engineering
University of Washington, Seattle, WA

Abstract

Computer vision software is complex, involving many tens of thousands of lines of code. Coding mistakes are not uncommon. When a vision algorithm is run on controlled data which meet all the algorithm assumptions, the results are often statistically predictable. This renders it possible to statistically validate the algorithm and its associated theoretical derivations. In this paper we review the general theory of some relevant kinds of statistical tests and then illustrate the experimental methodology of statistical algorithm validation to validate a program that estimates parameters of buildings in aerial photographs. This program estimates the 3D positions of building vertices based on input data obtained from multi-image photogrammetric resection calculations and 3D geometric information relating some of the points, lines and planes of the building to each other.

Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **FEB 2001**	2. REPORT TYPE	3. DATES COVERED **00-02-2001 to 00-02-2001**

4. TITLE AND SUBTITLE **Error Propagation and Statistical Validation of Computer Vision Software**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Language and Media Processing Laboratory,Institute for Advanced Computer Studies,University of Maryland,College Park,MD,20742-3275**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES **The original document contains color images.**

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**23**	

Error Propagation and Statistical Validation of Computer Vision Software

Xufei Liu, Tapas Kanungo, and Robert M. Haralick

Error Propagation and Statistical Validation
of Computer Vision Software

Xufei Liu,[1] Tapas Kanungo,[2] and Robert M. Haralick[3]

[1]Ranch Networks
Moganville, NJ 07751

[2]Center for Automation Research
University of Maryland, College Park, MD

[3]Dept. of Electrical Engineering
University of Washington, Seattle, WA

Abstract

Computer vision software is complex, involving many tens of thousands of lines of code. Coding mistakes are not uncommon. When a vision algorithm is run on controlled data which meet all the algorithm assumptions, the results are often statistically predictable. This renders it possible to statistically validate the algorithm and its associated theoretical derivations. In this paper we review the general theory of some relevant kinds of statistical tests and then illustrate the experimental methodology of statistical algorithm validation to validate a program that estimates parameters of buildings in aerial photographs. This program estimates the 3D positions of building vertices based on input data obtained from multi-image photogrammetric resection calculations and 3D geometric information relating some of the points, lines and planes of the building to each other.

1 Introduction

Many computer vision problems can be posed as either parameter estimation problems (for example, estimate the pose of an object) or hypothesis testing problems (for example, which of N objects in a database occurs in a given image.) Since the input data (such as images or feature points) is noisy, the estimates produced by computer vision algorithms are noisy. In other words, there is an inherent uncertainty associated with the results produced by any computer vision algorithm. These uncertainties are best expressed in terms of statistical distributions and their means and covariances. Details of the theory and application of covariance propagation can be found in [5, 8], and in the references cited in [8].

Usually, implementations of vision algorithms involve thousands of lines of code. Furthermore, the algorithms are often based on many approximations and mathematical calculations. One way to check whether the implementation and the theoretical calculations are correct is by providing the algorithm with input data with known (controlled) statistical characteristics (which is possible since the input data can be artificially generated) and then checking if the distribution of the output is in agreement with what was predicted by theoretical calculations.

Since many of the estimation problems in computer vision are multidimensional, testing whether the means and covariances of the empirical distribution and predicted distribution are the same is easier than testing whether the shapes of the two distributions are the same. In this paper, we summarize statistical tests for the case in which the estimates can be assumed to be multivariate Gaussian. We also describe the interfaces to software that we have implemented for conducting these tests. Although software libraries and environments (e.g. Splus, numerical recipes) are available for conducting such tests for one-dimensional samples, we are unaware of similar software libraries for the multivariate case. In fact, most statistics books do not give all five of the tests we give (e.g., Koch [12] does not address the fifth testing problem). The hypothesis testing theory and software are described in [10], and the software can be obtained at no cost from the Statlib software repository or from kanungo@cfar.umd.edu. A description of how the software and the theory have been tested using statistical techniques is also included. A preliminary version of this work was presented at the 1996 DARPA Image Understanding Workshop [16].

2 Kinds of Statistical Hypotheses

Let $x_1, x_2, \ldots, x_n$ be a sample from a multivariate Gaussian distribution with population mean μ and population covariance Σ. That is, $x_i \in R^p$ and $x_i \sim N(\mu, \Sigma)$, where p is the dimension of the vectors x_i.

We can formulate various hypotheses about the population mean and covariance, depending on what is known and what is unknown. The data x_i can then be used to test whether a hypothesis is false. Notice that each population parameter (here we have two, μ and Σ) can be either tested, or unknown and untested, or known. If a parameter is being tested, a claim about its value is being made. If a parameter is unknown and untested, no claim is being made about its value; its value is not known and therefore we

cannot use it in any computation. If the value of a parameter is assumed to be known, then we assume that its value is known without error and cannot be questioned or tested, just as the normality assumption is not questioned. Furthermore, when a parameter value is known, that value can be used in computations of test statistics for other parameters.

In general, if a distribution has q parameters, there can be $3^q - 2^q$ tests. The reasoning is as follows. Since each parameter can be either tested, or unknown and untested, or known, the number of possibilities is 3^q. But of these, the number of combinations in which none of the parameters are tested (that is, they are either known, or unknown and untested and so do not represent a test) is 2^q. Thus the total number of distinct hypotheses that can be made about a sample from a q-parameter distribution is $3^q - 2^q$.

In the case when the data come from a multivariate normal distribution, $N(\mu, \Sigma)$, we have $q = 2$ and thus can have $3^2 - 2^2 = 5$ possible hypotheses. The five tests are as follows:

H_1 : $\mu = \mu_0$ ($\Sigma = \Sigma_1$ **known.**) In this test, the question is whether or not the sample is from a Gaussian population whose mean is μ_0. The population covariance Σ is assumed to be known and equal to Σ_1.

H_2 : $\mu = \mu_0$ (Σ **unknown, untested.**) In this test, the question is whether or not the sample is from a Gaussian population whose mean is μ_0. No statement is made regarding the population covariance Σ.

H_3 : $\Sigma = \Sigma_0$ ($\mu = \mu_1$ **known.**) In this test, the question is whether or not the sample is from a Gaussian population whose covariance is Σ_0. The population mean μ is assumed to be known and equal to μ_1.

H_4 : $\Sigma = \Sigma_0$ (μ **unknown, untested.**) In this test, the question is whether or not the sample is from a Gaussian population whose covariance is Σ_0. No statement is made regarding the mean μ.

H_5 : $\mu = \mu_0$ $\Sigma = \Sigma_0$ In this test, the question is whether or not the sample is from a Gaussian population whose mean is μ_0 and whose covariance is Σ_0. This is the principal test we use for software validation.

3   Definitions

In this section we briefly define the terms used in the rest of the paper. A reader who is familiar with statistics can skip this section. For a lucid explanation of the basic univariate concepts see [4]. A slightly more rigorous treatment of univariate and multivariate tests is given in [2]. Multivariate tests are treated in great detail in [12]. The most authoritative reference on multivariate statistics is [1]. Although this book has most of the results given here, it is not easy to read, and the results are not easy to find in it.

A *statistic* of the data $x_1, \ldots, x_n$ is any function of the data. For example, the sample mean $\bar{x}$ is a statistic, and so is the sample covariance matrix S. The statistic need not be one-dimensional; $(\bar{x}, S)^t$ together are another statistic of the same data. A *sufficient*

2

statistic is a statistic that contains all the information about the data; any inference regarding the underlying population can be made using just the sufficient statistic, and the individual data points do not add any more information to the inference process. For example, the vector of original data $(x_1, \ldots, x_n)^t$ is a sufficient statistic; it contains all the information about the data. Another sufficient statistic is $(\bar{x}, S)^t$. A sufficient statistic is not unique. A *minimal sufficient statistic* is a sufficient statistic that has a smallest number of entries. For example, for Gaussian data, $(\bar{x}, S)$ is a minimal sufficient statistic.

A *hypothesis* is any statement about a population parameter that is either true or false. The *null hypothesis* H_0 and the *alternate hypothesis* H_A are the two complementary hypotheses in a statistical hypotheses testing problem.

A *test statistic* is a statistic of the data that is used for testing a hypothesis. The *null distribution* is the distribution of the test statistic when the null hypothesis is true. The *alternate distribution* is the distribution of the test statistic when the alternate hypothesis is true.

There are two types of errors — mis-detection and false alarm. If the null hypothesis is true but the test procedure decides it is false, this is called a *misdetection.* When the alternate hypothesis is true but the test procedure accepts the null hypothesis, this is called a *false alarm.* The misdetection probability of a test procedure is usually fixed by the user; this is referred to as the *significance level,* α, of the test. A typical value for α is 0.05.

The *power function* of a hypothesis test is a function of the population parameter θ; its value, $\beta(\theta)$, is equal to 1 minus the probability of false alarm. Ideally, the power function should be 0 for all θ such that the null hypothesis is true, and 1 for all θ such that the alternate hypothesis is true. For most realistic testing problems one cannot create a test procedure with such an ideal power function. Power functions are very useful for evaluating hypothesis testing procedures. For an example in which they are used for computer vision problems see [11]. A *uniformly most powerful test* is a test procedure whose power function is higher than those of all other tests.

There are many methods of creating tests and corresponding test statistics. The test statistics given in this paper were derived by maximizing the likelihood ratio. The derivation of these statistics can be found in the cited literature.

4   Test statistics

In this section we summarize various test statistics and their distributions under the true null hypothesis For details see Anderson [1] and Kanungo and Haralick [10]. These tests will be used later in our discussion. We use the following definitions of $\bar{x}$ and S :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^t,$$

3

where we have assumed that the data vectors x_i are p-dimensional and the sample size is n.

4.1   Test 1: $\mu = \mu_0$ with known $\Sigma = \Sigma_1$

Test statistic:
$$T = n(\bar{x} - \mu_0)^t \Sigma_1^{-1}(\bar{x} - \mu_0). \tag{1}$$

Its distribution under the null hypothesis is Chi-squared:
$$T \sim \chi_p^2.$$

The alternate hypothesis is $H_A : \mu \neq \mu_0$; the distribution of the test statistic under the alternate hypothesis is non-central Chi-squared:
$$T \sim \chi_{p,d}^2$$

where $d = n(\mu - \mu_0)^t \Sigma_1^{-1}(\mu - \mu_0)$ is the non-centrality parameter.

4.2   Test 2: $\mu = \mu_0$ with unknown Σ

Hotelling's Test statistic:
$$T = \frac{n(n - p)}{p(n - 1)}(\bar{x} - \mu_0)^t S^{-1}(\bar{x} - \mu_0). \tag{2}$$

Its distribution under the null hypothesis is F:
$$T \sim F_{p,n-p}.$$

The alternate hypothesis is $H_A : \mu \neq \mu_0$; the distribution of the test statistic under the alternate hypothesis is non-central F:
$$T \sim F_{p,n-p,d}$$

where $d = n(\mu - \mu_0)^t \Sigma^{-1}(\mu - \mu_0)$ is the non-centrality parameter.

4.3   Test 3: $\Sigma = \Sigma_0$ with known $\mu = \mu_1$

Let
$$C = \sum_{i=1}^{n}(x_i - \mu_1)(x_i - \mu_1)^t = (n - 1)S + (\bar{x} - \mu_1)(\bar{x} - \mu_1)^t .$$

and
$$\lambda = (e/n)^{pn/2}|C\Sigma_0^{-1}|^{n/2} \exp(-tr(C\Sigma_0^{-1})/2) .$$

The test statistic is
$$T = -2 \log \lambda. \tag{3}$$

Its distribution under the null hypothesis is Chi-squared:
$$T \sim \chi_{p(p+1)/2}^2.$$

The alternate hypothesis is $H_A : \Sigma \neq \Sigma_0$; the distribution of the test statistic under the alternate hypothesis is unknown.

4

4.4 Test 4: $\Sigma = \Sigma_0$ with unknown μ

Let $B = (n-1)S$, and $\lambda = (e/(n-1))^{p(n-1)/2} |B\Sigma_0^{-1}|^{(n-1)/2} \exp(-tr(B\Sigma_0^{-1})/2)$ The test statistic is

$$T = -2\log\lambda. \tag{4}$$

Its distribution under the true hypothesis is Chi-squared:

$$T \sim \chi^2_{p(p+1)/2}.$$

The alternate hypothesis is $H_A : \Sigma \neq \Sigma_0$; the distribution of the test statistic under the alternate hypothesis is unknown.

4.5 Test 5: $\Sigma = \Sigma_0$ and $\mu = \mu_0$

Define $B = (n-1)S$ and

$$\begin{aligned}
\lambda &= (e/n)^{pn/2} |B\Sigma_0^{-1}|^{n/2} \\
&\quad \exp\left(-[tr(B\Sigma_0^{-1}) + n(\bar{x} - \mu_0)^t \Sigma_0^{-1}(\bar{x} - \mu_0)]/2\right)
\end{aligned}$$

The test statistic is

$$T = -2\log\lambda \tag{5}$$

Its distribution under the null hypothesis is Chi-squared:

$$T \sim \chi^2_{p(p+1)/2+p}$$

The alternate hypothesis is $H_A : \Sigma \neq \Sigma_0$, and $\mu \neq \mu_0$; the distribution of the test statistic under the alternate hypothesis is unknown.

5 Validating theory and software

To validate computer vision software two checks have to be performed. The first check is that the theory is correct: the theoretically derived null distributions of the test statistics are actually correct. The second check is that the software is correct: the implementation is exactly what the theory dictates. Both of these checks can be done by computing the empirical distributions and comparing them with the theoretically derived distributions. In the next subsection we describe how we empirically compute the null distributions of the five test statistics, and in the following section we describe how we use the Kolmogorov-Smirnov test to check if the empirical distribution and the theoretically-derived distributions are the same.

5.1 Empirical null distributions

In order to generate the empirical null distributions we proceed as follows.

1. Choose some values for the multivariate Gaussian population parameters p, μ and Σ.

2. Generate n samples from the population.

3. Compute the value of the statistic T for the test that is being verified.

4. Repeat steps 2 and 3 M times to get T_i, $i = 1, \ldots, M$.

5. The empirical distribution T can be computed by computing the histogram of T_i.

5.2 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov (KS) test tests whether two distributions are alike. The KS test uses the fact that the maximum absolute difference between the empirical cumulative distribution (the KS test statistic) and the theoretical cumulative distribution has a known distribution (the null distribution). For a more detailed discussion of the KS test see [17].

6 Application: 3D Parameter Estimation

We applied our hypothesis testing methodology to validate the 3D parameter estimation software that was used for constructing the ground truth model from the RADIUS model board data set [18]. In this section we describe the problem and the optimization approach.

6.1 Site Model Construction

The task is to construct 3D object models from detected 2D image features and the known geometric constraints on the observed perspective projections of the 3D objects. The data set [18] consists of 78 images of the two RADIUS model boards and the 3D coordinates of some building vertices. Since the purpose was to establish ground truth for automatic site model construction algorithms, the corresponding points of the building vertices that were observable on the images were identified and located manually. Also, the 3D positions of a few of the building vertices are known. Simultaneous estimation of the interior parameters and exterior orientation parameters of the cameras was done by setting up and solving a very large photogrammetric resection problem. Then, using these camera parameters, multi-image triangulation was performed. This yielded noisy estimates for the building vertices; these estimates were the input to the site model construction software whose testing we now describe.

The geometric constraint procedure takes the photogrammetrically estimated 3D point positions and their covariance matrices as observations. It uses the partial models of the buildings to generate constraints on the building parameters. To estimate the optimal 3D parameters that satisfy the relations in the partial models, a constrained optimization model is solved. By error propagation we derive the covariance matrix of the estimated building vertices, which are now guaranteed to satisfy the given constraints. This is discussed in greater detail in [15].

6.2 Constrained Optimization

Partial building models are used to constrain the 3-D parameters to be estimated. Partial models represent geometric constraints between building entities: lines, planes and vertices. The specific relations we used to constrain the parameters were: point-on-plane, point-on-line, plane-angle-plane, line-angle-line, and plane-angle-line. For a more detailed description see [13].

The observed 3D points and the associated covariance matrix Σ are obtained by triangulation. The perturbation model that we used for the observations was zero-mean Gaussian noise with unknown covariance. Having the partial object model and the perturbation model, we can define the estimation problem. Let $\Theta \in I\!\!R^m$ denote the parameters, $X' \in I\!\!R^m$ the observations, and $p(X' \mid \Theta)$ the likelihood function. In the building estimation problem, the parameters are the coordinates of the points, the normal vectors and distance constants of the planes, and the direction cosines and reference points of the lines.

Assuming that the optimality criterion is the maximum posterior probability, a Bayesian approach can be used to transform the problem into a maximum likelihood problem with constraints. Let the constraints be denoted by $\Theta \in C_\Theta \subset I\!\!R^m$. This problem can be expressed as a constrained optimization problem.

$$\min\left\{-p(X' \mid \Theta) \mid \Theta \in C_\Theta\right\}$$

The problem can be reformulated by taking the logarithm of the probability function. Under the assumption of Gaussian noise, we obtain a least squares model. The objective function is the sum of squared errors between the estimated point positions and the observed points.

$$\min_{\Theta} \quad \left\{f(\Theta) := (X' - X)^T \Sigma^{-1} (X' - X)\right\} \tag{6}$$
$$\text{subject to} \quad \Theta \in C_\Theta$$

where X denotes the unknown 3D points, and the feasible set C_Θ is determined by the partial model and the unit length constraint on the directional vectors.

If the noise affecting different 3D points is independent, the objective function can be rewritten as

$$f(\Theta) = \sum_{i=1}^{K} (\mathbf{x}_i' - \mathbf{x}_i)^T \Sigma_i^{-1} (\mathbf{x}_i' - \mathbf{x}_i)$$

where Σ_i is the covariance matrix of the ith point and K is the number of observed points.

The constraints can be incorporated into the optimization problem as follows (see [13]):

$$\min_{\Theta} \quad \left\{f(\Theta) := (X' - X)^T \Sigma^{-1} (X' - X)\right\} \tag{7}$$
$$\text{subject to} \quad h_i(\Theta) = 0, \quad i = 1, ..., r$$

In the above, the equation $h_i(\Theta) = 0$ represents all the constraints derived from the partial models. The numerical solution to this optimization problem can be achieved by various methods such as: the reduced gradient method [7], sequential quadratic programming [6], or the augmented Lagrangian method [3].

7 Error Propagation

Once the constrained optimization produces a result, we use the error propagation approach [9] [14] to transform the input error covariance matrix to the output covariance matrix. In the building estimation problem, we have the optimization model

$$\min_{\theta} \quad f(\Theta)$$
$$\text{subject to} \quad \mathbf{h}(\Theta) = 0$$

where f is the sum of squared errors between the estimated 3D points and the observed 3D points.

The Lagrangian function is

$$L(X', \Theta, \Lambda) = f(X', \Theta) + \Lambda^T \mathbf{h}(\Theta)$$

Suppose $(\tilde{X}, \tilde{\Theta}, \tilde{\Lambda})$ is an optimal point. From the necessary conditions on a local minimum point, the linearized model at the optimal point can be obtained [13, 6] by solving

$$\left(\begin{array}{cc} \tilde{\mathbf{Q}} & \tilde{\mathbf{H}}^T \\ \tilde{\mathbf{H}} & 0 \end{array} \right) \left(\begin{array}{c} \Delta\Theta \\ \Delta\Lambda \end{array} \right) = \left(\begin{array}{c} -\tilde{\mathbf{B}}\Delta X \\ 0 \end{array} \right). \tag{8}$$

The Lagrangian matrix at the point of $(\tilde{X}, \tilde{\Theta}, \tilde{\Lambda})$ can be approximated by the Lagrangian matrix at the minimum if the error is small. Hence the linear model can be approximated by

$$\left(\begin{array}{cc} \mathbf{Q}^* & (\mathbf{H}^*)^T \\ \mathbf{H}^* & 0 \end{array} \right) \left(\begin{array}{c} \Delta\Theta \\ \Delta\Lambda \end{array} \right) = \left(\begin{array}{c} -\mathbf{B}^*\Delta X \\ 0 \end{array} \right)$$

where

$$
\begin{aligned}
\mathbf{Q}^* &= \nabla^2_{\theta\theta} \mathcal{L}(X', \hat{\Theta}, \hat{\Lambda}) \\
&= \nabla^2 f(X', \hat{\Theta}) + \sum_{j=1}^{r} \hat{\lambda}_j \nabla^2 h_j(\hat{\Theta}) \\
\mathbf{B}^* &= \nabla^2_{\theta\mathbf{x}} \mathcal{L}(X', \hat{\Theta}, \hat{\Lambda}) = \nabla^2_{\theta\mathbf{x}} f(X', \hat{\Theta}) \\
\mathbf{H}^* &= \nabla \mathbf{h}(\hat{\Theta})
\end{aligned}
$$

Assume that the constraints are linearly independent. Then the row vectors in matrix H^* are linearly independent. We can use the null space method to compute the error propagation matrix J [6] [13].

Once the error propagation matrix is obtained, we can propagate the covariance matrix of the observations Σ to the output. The covariance matrix of the estimated parameters Σ_Θ can be approximated by

$$\Sigma_\Theta = J\Sigma J^T \tag{9}$$

8

8 Experimental Methodology

To validate the optimization algorithm and the error propagation model, an experiment is needed. This section describes the experimental methodology for this validation.

8.1 Ideal Data Generators and Noise Model

Three building types, the cube box, the peak roof house, and the hip roof house, appear frequently in the given sites. The peak roof house model assumes a slanted roof on a cube. The hip roof house model assumes a roof with slanted sides and ends. For a more detailed description of the models and corresponding constraints see [13]. They are chosen as our prototype models with unknown length, location and orientation parameters.

In our experiment, ideal data generators randomly generate the ideal parameters for the prototype models and produce ideal 3D points.

Assume that a 3D coordinate system x-y-z is used. To simulate the site model situation, the ground is assigned as the plane $z = 0$. Without losing generality we assume that the ideal model parameters that determine the 3D positions of the building vertices have uniform distributions. The center of the bottom plane of a basic model is in a region defined by $[-x_0, x_0), [-y_0, y_0), [-z_0, z_0)$. The basic model is rotated on the ground by a random angle $\phi \in [\phi_0, \phi_1)$.

The three length parameters for the cube box model are denoted by a, b, c, with $a_0 \leq a < a_1, b_0 \leq b < b_1$, and $c_0 \leq c < c_1$. In our experiment the ranges of the parameters for this model were set as follows:

x_0	y_0	z_0	ϕ_0	ϕ_1	a_0	a_1	b_0	b_1	c_0	c_1
50	50	0	0	2π	30	60	30	60	30	60

The peak roof model uses four length parameters, a, b, c, d, where a, b, c are the same as those in cube model and their ranges are the same. The height of the peak roof is defined by a parameter d, with $10 \leq d \leq 20$. The hip roof model requires one more parameter e; the length of the roof edge is $a - 2e$, with $5 \leq e \leq 10$.

For each building type, K ideal buildings are randomly generated. Each of these K buildings will be used in n experiments in which Gaussian random noise is added to each of the 3D coordinates of the building and constrained optimization is used to estimate the building vertices that satisfy the geometric constraints. As a result of these n experiments, n estimates of the building parameters are produced. It is these n estimates on which the hypothesis test statistics will be computed. We call the procedure for determining these n test statistics a trial. Since there are K ideal buildings for each ideal building type, we can compute K test statistics. These K statistics can then be used to test the hypothesis that their distribution is what the statistical theory of the test says it should be.

The noise values are independently sampled from a Gaussian distribution $\mathcal{N}(0, \sigma^2 I)$, where σ is the standard deviation of the random variables $\delta x, \delta y, \delta z$. We repeated each experiment with σ set to 1.0, 2.0 or 3.0. The validation results for all three different standard deviations are similar, so here we discuss only the validation for the case where

9

the standard deviation is equal to 3.0. K is set at 100. n is set at 500 for cube model
and 700 for other models.

8.2 Statistical Test

In each trial a sample of model parameters and corresponding ideal 3D points s produced
by the ideal data generator. Let the ideal parameters be denoted by $\tilde{\Theta}$. For each ideal
building instance having parameters $\tilde{\Theta}$, n independent perturbations $\{\Delta X_i, \ i = 1, ..., n\}$
are generated from the noise model with distribution $\mathcal{N}(0, \Sigma)$. By adding the pertur-
bations to the ideal points, the perturbed data set $\{X'_1, X'_2, ..., X'_n\}$ is generated. For
each of the perturbed data sets $\{X'_1, X'_2, ..., X'_n\}$, an optimal solution $\hat{\Theta}_i$ is computed by
solving

$$\min_{\Theta_i} \quad f(X'_i, \Theta_i) \qquad i = 1, ..., n$$
$$\text{subject to} \quad \mathbf{h}(\Theta_i) = 0$$

Thus we have n estimates $\{\hat{\Theta}_i, \ i = 1, ..., n\}$.

Using equation (9), we can transform the input covariance matrix through the error
propagation matrix to the output. If the linear model is valid, the estimated parameters
$\{\hat{\Theta}, \ i = 1, ..., n\}$ should be approximately distributed as $\mathcal{N}(\tilde{\Theta}, J\Sigma J^T)$.

Let $\Delta \hat{\Theta}_i$ denote $\hat{\Theta}_i - \tilde{\Theta}$, $i = 1, ..., n$. Let $\mu_0 = 0$ and $\Sigma_0 = J\Sigma J^T$. Under the linearized
model, $\{\Delta \hat{\Theta}_i, \ i = 1, ..., n\}$ have distribution $\mathcal{N}(\mu_0, \Sigma_0)$. Considering $\{\Delta \hat{\Theta}_i, \ i = 1, ..., n\}$
as a random sample from a Gaussian distribution $\mathcal{N}(\mu, \Sigma)$, we can perform any one of
the five hypothesis tests. Here we just discuss our results for hypothesis H_5: $\mu = \mu_0$ and
$\Sigma = \Sigma_0$. The results for the other hypothesis tests are similar.

The significance level α is selected to be 0.05. Under the null hypothesis, the computed
statistics of the mean and covariance tests have the null distributions. This can be verified
by using a Kolmogorov-Smirnov test (K-S test) on the K test statistics generated from
the K trials.

8.3 Range Space Analysis

The standard hypothesis test methods require that the covariance matrix be positive
definite. However, because of the constraints, the output covariance of a constrained
optimization is generally positive semi-definite.

Theorem 1 *Suppose that not all of the derivatives of the constraint equations are equal
to zero at the local minimum point, then the propagated error covariance $J\Sigma J^T$ is singular.*

<u>Proof</u> : From the given condition we know that the derivative matrix H is not a zero
matrix, i.e.,

$$H = \begin{pmatrix} \frac{\partial h_1}{\partial \Theta} \\ . \\ . \\ . \\ \frac{\partial h_r}{\partial \Theta} \end{pmatrix} \neq 0$$

10

Left multiply equation $\Delta\Theta = J\Delta X$ by H :

$$H\Delta\Theta = HJ\Delta X$$

Because both (X, Θ^*, Λ^*) and $(X + \Delta X, \Theta^* + \Delta, \Lambda^* + \Delta\Lambda)$ are local minimum points of the optimization, the following equation is satisfied,

$$H\Delta\Theta = 0.$$

Hence

$$0 = HJ\Delta X$$

Since the formula holds for any ΔX, this implies that

$$0 = HJ \tag{10}$$

We now use this result to prove that $J\Sigma J^T$ is singular. Left multiply $J\Sigma J^T$ by H and right multiply it by H^T. From (10) we have

$$HJ\Sigma J^T H^T = 0\Sigma 0$$

Since H is not a zero matrix, $J\Sigma J^T$ must be singular. $\square$

To utilize standard hypothesis technology, we project a positive semi-definite matrix onto its range space. Suppose that an $n \times n$ covariance matrix Σ_0 has k nonzero eigenvalues $w_1, ..., w_k$ and associated unit eigenvectors $v_1, ..., v_k$. A basis of the range space of Σ_0 can be defined by

$$B = (v_1, ..., v_n)$$

Use B to perform a matrix transformation as follows:

$$B^T \Sigma_0 B = \Sigma_B = \begin{pmatrix} w_1 & 0 & . & . & . \\ 0 & w_2 & 0 & . & . \\ . & & . & . & \\ . & & . & . & \\ 0 & . & . & . & w_k \end{pmatrix}$$

Let $B^\perp$ be a basis matrix of the null space of matrix Σ_0. It is obvious that $(B, B^\perp)$ is orthonormal. In our experiment, we check whether $(B^\perp)^T \Delta\Theta$ has very small variances (caused by round-off errors and nonlinear terms). If this is true, we conduct the hypothesis test on variables $B^T \Delta\Theta$ with covariance matrix Σ_B.

Due to the round-off errors and nonlinear terms, the zero eigenvalues of matrix Σ_0 may not be exactly zero. We use a small threshold to distinguish the zero eigenvalues from the nonzero eigenvalues. In all our experiments the threshold was set to 10^{-6} times the maximum eigenvalue.

For the cube model, the range space of the output error covariance matrix has 7 dimensions. This can be understood as follows. Consider a cube house whose faces are all at right angles to each other. Count the number of degrees of freedom. The size of a cube model is defined by three independent parameters. The location of the model is

11

specified by three translation parameters in 3D space. In our experiments, the normal vector of the cube roof is fixed in the vertical direction. The only possible rotation is around the vertical axis of the model. Thus the total number of independent parameters is seven. For the peak roof model the analysis is similar, except that two more parameters are needed to determine the roof height and the ridge position. (In the partial model we do not fix the horizontal position of the roof ridge to the center of the building.) Thus the range space of the output covariance matrix for the peak roof model has nine dimensions. The hip roof model inherits all the parameters of the peak roof model. It requires two more parameters to determine how much of the ridge was cut of at each of the two ends (they are assumed to be independent). These parameters can be thought of as the relative starting and ending points of the the ridge. Thus the range space of the output covariance matrix for the hip roof model has 11 dimensions.

9   Experimental Results

In this section we present our experimental results on the cube model, the peak roof model and the hip roof model. An example of the model estimation results is shown in Figure 4.

9.1   Test of Cube Model with $\sigma = 3.0$

The theoretical and empirical null distributions of the five test statistics for the cube model are shown in Figure 1. The standard deviation used was $\sigma = 3.0$. The x axis is the statistic used in the test and the y axis represents $1 - \alpha$, where α is the significance level.

The experimental trials were run multiple times; each time, the null hypothesis was either rejected or not rejected. The results are summarized in Table 1(a). The null hypothesis is not rejected at a 0.05 significance level.

The K-S test was used to test whether the empirical and theoretical distributions are similar. For this test the number of degrees of freedom, p, is 7. The results are shown in Table 2(b). None of the five test statistic distributions failed the K-S test at a significance level of 0.05. Thus the optimization model and the error propagation model for the cube model were validated.

9.2   Test of Peak Roof Model with $\sigma = 3.0$

The theoretical and empirical null distribution of the five test statistics for the peak roof model are shown in Figure 2. The standard deviation used was $\sigma = 3.0$. The x axis is the statistic used in the test and the y axis represents $1 - \alpha$, where α is the significance level.

The experimental trials were run multiple times; each time, the null hypothesis was either rejected or not rejected. The results are summarized in Table 2(a). The null hypothesis is not rejected at a 0.05 significance level.

The K-S test was used to test whether the empirical and theoretical null distributions are similar. For this test the number of degrees of freedom, p, is 9. The results are
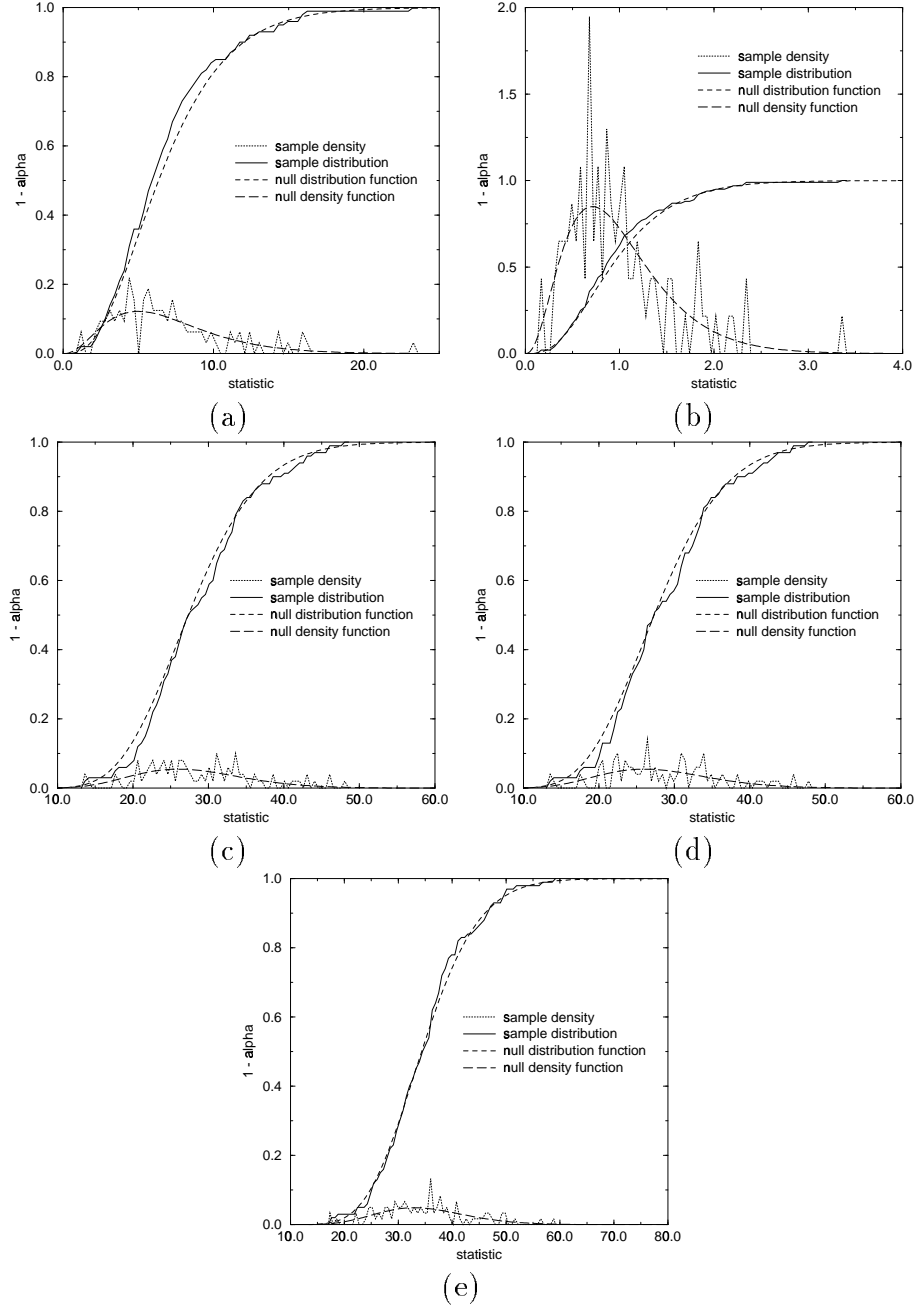
Figure 1: Null distributions of the five test statistics for the cube model at $\sigma = 3$. (a) Test mean with known covariance. (b) Test mean with unknown covariance. (c) Test covariance with known mean. (d) Test covariance with unknown mean. (e) Test mean and covariance.

Table 1: Cube model.

(a) Test Statistic Values for the Five Tests

test method	number of trials	sample size	significance level	reject rate
test μ, known Σ	100	500	0.05	0.06
test μ, unknown Σ	100	500	0.05	0.05
test Σ, known μ	100	500	0.05	0.07
test Σ, unknown μ	100	500	0.05	0.06
test μ and Σ	100	500	0.05	0.05

(b) Kolmogorov-Smirnov Distribution Test

test method	null distribution	size n	p-value	K-S test
test μ, known Σ	χ^2_p	100	0.391	pass
test μ, unk. Σ	$F_{p,n-p}$	100	0.217	pass
test Σ, known μ	$\chi^2_{p(p+1)/2}$	100	0.866	pass
test Σ, unk. μ	$\chi^2_{p(p+1)/2}$	100	0.881	pass
test μ and Σ	$\chi^2_{p(p+1)/2+p}$	100	0.700	pass

shown in Table 2(b). None of the five test statistic distributions failed the K-S test at a significance level of 0.05. Thus the optimization model and the error propagation model for the peak roof model were validated.

9.3   Test of Hip Roof Model with $\sigma = 3.0$

The theoretical and empirical null distributions of the five test statistics for the hip roof model are shown in Figure 3. The standard deviation used was $\sigma = 3.0$. The x axis is the statistic used in the test and the y axis represents $1 - \alpha$, where α is the significance level.

The experimental trials were run multiple times; each time, the null hypothesis was either rejected or not rejected. The results are summarized in Table 3(a). The null hypothesis was not rejected at a 0.05 significance level.

The K-S test was used to test whether the empirical and theoretical null distributions are similar. For this test the number of degrees of freedom, p, is 11. The results are shown in Table 3(b). None of the five test statistic distributions failed the K-S test at a significance level of 0.05. Thus the optimization model and the error propagation model for the hip roof model were validated.

10   Summary

We have described a statistical methodology for validating the theoretical derivations and software that make up a system for estimating 3D positions of building vertices, based on input data obtained from multi-image photogrammetric resection calculations. Error propagation allowed us to derive the null distributions of various test statistics of mea-
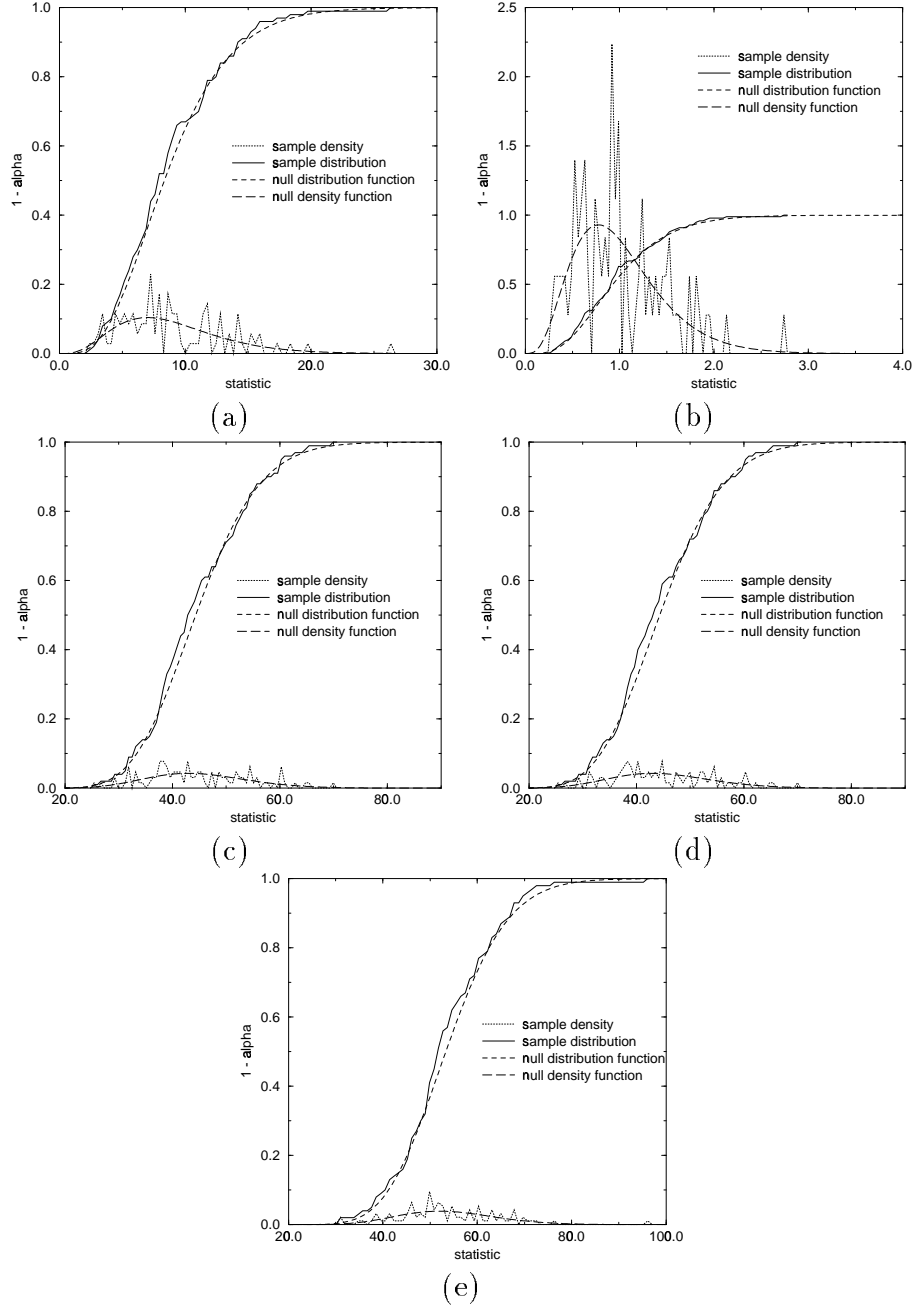
Figure 2: Null distributions of the five test statistics for the peak roof model at $\sigma = 3$. (a) Test mean with known covariance. (b) Test mean with unknown covariance. (c) Test covariance with known mean. (d) Test covariance with unknown mean. (e) Test mean and covariance.
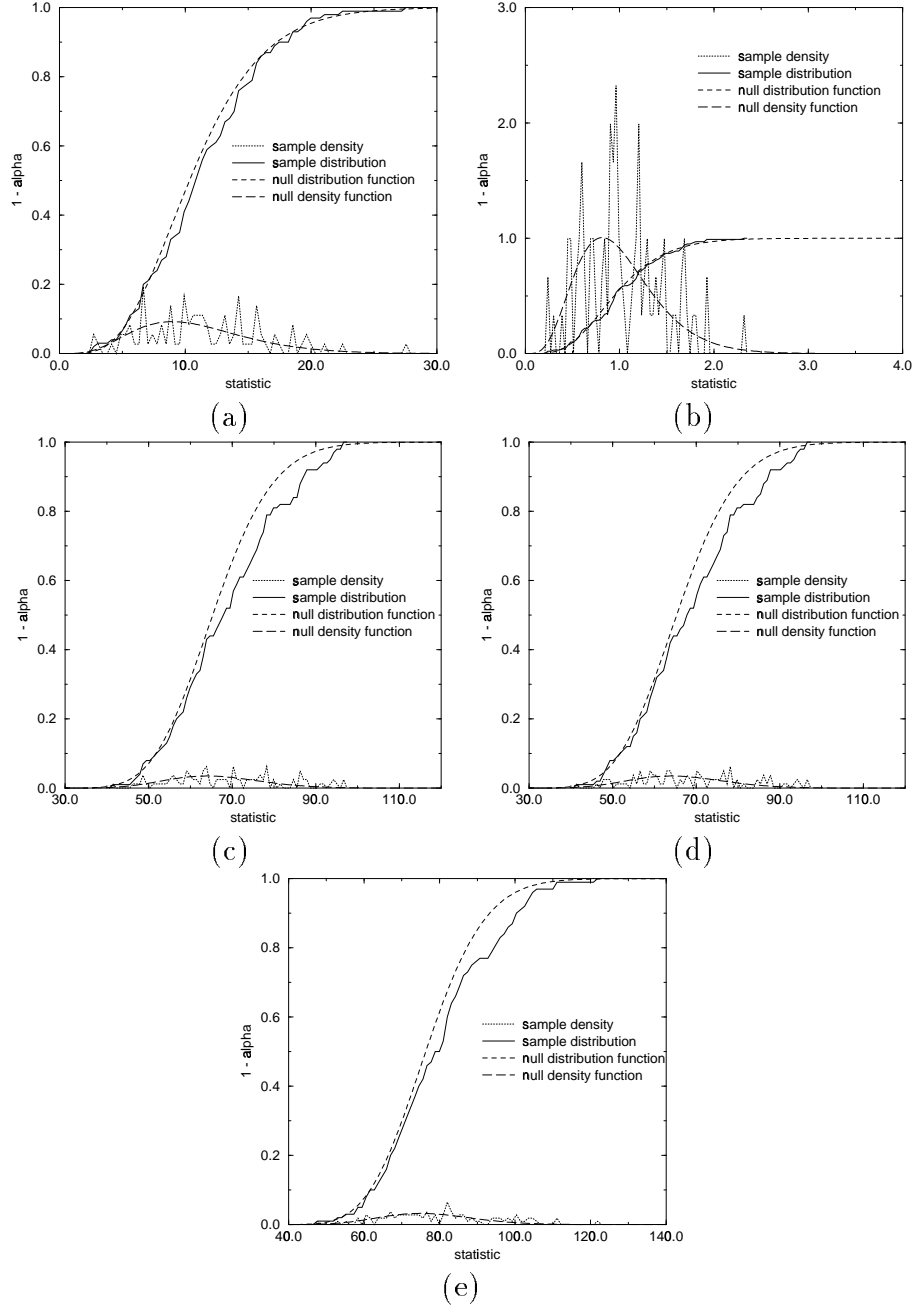
Figure 3: Null distributions of the five test statistics for the hip roof model at $\sigma = 3$. (a) Test mean with known covariance. (b) Test mean with unknown covariance. (c) Test covariance with known mean. (d) Test covariance with unknown mean. (e) Test mean and covariance.
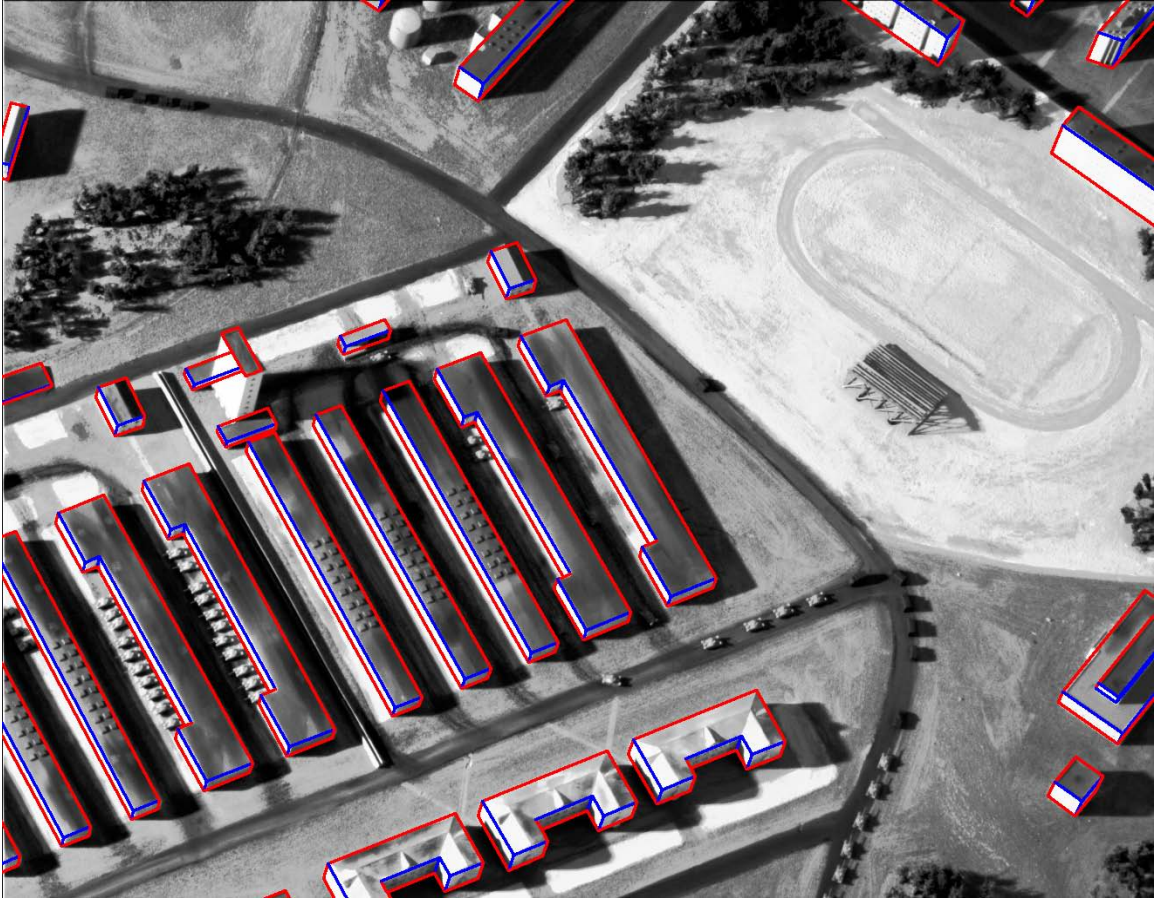
Figure 4: This figure shows the estimated houses overlaid on top of the original model-board image.

Table 2: Peak roof model results.

(a) Test Statistic Values for the Five Tests

test method	number of trials	sample size	significance level	reject rate
test μ, known Σ	100	700	0.05	0.03
test μ, unk. Σ	100	700	0.05	0.03
test Σ, known μ	100	700	0.05	0.04
test Σ, unk. μ	100	700	0.05	0.04
test μ and Σ	100	700	0.05	0.02

(b) Kolmogorov-Smirnov Distribution Test

test method	null distribution	size n	p-value	K-S test
test μ, known Σ	χ^2_p	100	0.464509	pass
test μ, unknown Σ	$F_{p,n-p}$	100	0.277997	pass
test Σ, known μ	$\chi^2_{p(p+1)/2}$	100	0.450701	pass
test Σ, unknown μ	$\chi^2_{p(p+1)/2}$	100	0.402053	pass
test μ and Σ	$\chi^2_{p(p+1)/2+p}$	100	0.205762	pass

surable quantities. These theoretically derived null distributions allowed us to validate whether the measurements in the system followed the theoretically derived distributions. Kolmogorov-Smirnov tests were performed to check whether the empirical and theoretical distributions were close enough. None of the empirically computed null distributions failed the KS test. Thus we have confirmed that the theoretical derivations of the null distributions were correct and that the software implementing the theory is also correct. The software for performing the statistical tests has been made publicly available.

Acknowledgment

References

[1] T. W. Anderson. *Introduction to Multivariate Statistical Analysis*. Wiley, New York, 1984.

18

Table 3: Hip roof model.

(a) Test Statistic Values for the Five Tests

test method	number of trials	sample size	significance level	reject rate
test μ, known Σ	100	700	0.05	0.03
test μ, unknown Σ	100	700	0.05	0.03
test Σ, known μ	100	700	0.05	0.11
test Σ, unknown μ	100	700	0.05	0.11
test μ and Σ	100	700	0.05	0.13

(b) Kolmogorov-Smirnov Distribution Test

test method	null distribution	size n	p-value	K-S test
test μ, known Σ	χ^2_p	100	0.771912	pass
test μ, unknown Σ	$F_{p,n-p}$	100	0.588617	pass
test Σ, known μ	$\chi^2_{p(p+1)/2}$	100	0.110815	pass
test Σ, unknown μ	$\chi^2_{p(p+1)/2}$	100	0.103598	pass
test μ and Σ	$\chi^2_{p(p+1)/2+p}$	100	0.096341	pass

[2] S. F. Arnold. *Mathematical Statistics*. Prentice-Hall, Englewood Cliffs, 1990.

[3] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982.

[4] G. Casella and R. L. Berger. *Statistical Inference*. Wadsworth, Belmont, CA, 1990.

[5] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, MA, 1993. Section 5.6.

[6] R. Fletcher. *Practical Methods of Optimization*. Wiley, New York, 1987.

[7] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Academic Press, New York, 1981.

[8] R. M. Haralick. Propagating covariance in computer vision. In *Proc. of IAPR Intl. Conf. on Pattern Recognition*, pages 493–498, Jerusalem, Israel, October 1994.

[9] D. A. Hudson. Inverse perspective projection algorithm: Error estimation. Master's thesis, University of Washington, Seattle, WA, 1990.

[10] T. Kanungo and R. M. Haralick. Multivariate hypothesis testing for Gaussian data: Theory and software. Technical Report ISL-TR-95-05, University of Washington, Seattle, WA, October 1995. Software available from Statlib:http://lib.stat.cmu.edu/ general/mvnetest.gz.

[11] T. Kanungo, R. M. Haralick, and H. S. Baird. Power functions and their use in selecting distance functions for document degradation model validation. In *Proc. of*

IAPR Intl. Conf. on Document Anaysis and Recognition, Montreal, Canada, August 1995.

[12] K. Koch. *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer-Verlag, New York, 1987.

[13] X. Liu. *Modeling and performance characterization of 3D parameter estimation using perspective geometry*. PhD thesis, University of Washington, Seattle, WA, 1995.

[14] X. Liu, R. M. Haralick, and A. Bedekar. Optimization methods for estimating 3D object parameters. In *Proc. of DARPA Image Understanding Workshop*, pages 1017–1025, Monterey, CA, 1994.

[15] X. Liu, R. M. Haralick, and K. Thornton. Site model construction using geometric constrained optimization. In *Proc. of DARPA Image Understanding Workshop*, Palm Springs, CA, 1996.

[16] X. Liu, T. Kanungo, and R. M. Haralick. Statistical validation of computer vision software. In *Proc. of DARPA Image Understanding Workshop*, pages 1533–1540, Palm Springs, CA, 1996.

[17] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, New York, 1990.

[18] K. Thornton, D. C. Nadadur, V. Ramesh, X. Liu, X. Zhang, A. Bedekar, and R. M. Haralick. Groundtruthing the RADIUS model-board imagery. In *Proc. of DARPA Image Understanding Workshop*, pages 319–329, San Francisco, CA, 1994.